Structure, Vol. 12, 1015–1025, June, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.str.2004.03.023

The Three-Dimensional Structure and X-Ray Sequence Reveal that Trichomaglin Is a Novel S-like Ribonuclease

Jian-Hua Gan,[1,4] Lu Yu,[2] Jian Wu,[1,5] Hong Xu,[3] Joyti S. Choudhary,[2] Walter P. Blackstock,[2] Wang-Yi Liu,[3] and Zong-Xiang Xia[1,*]

[1]State Key Laboratory of Bio-organic and Natural Products Chemistry
Shanghai Institute of Organic Chemistry
Chinese Academy of Sciences
Shanghai 200032
China

[2]Cell Map Project
GlaxoSmithKline
Stevenage
United Kingdom

[3]State Key Laboratory of Molecular Biology
Institute of Biochemistry and Cell Biology
Shanghai Institutes for Biological Sciences
Chinese Academy of Sciences
Shanghai 200031
China

Summary

Trichomaglin is a protein isolated from root tuber of the plant Maganlin (*Trichosanthes Lepiniate*, Cucurbitaceae). The crystal structure of trichomaglin has been determined by multiple-isomorphous replacement and refined at 2.2 Å resolution. The X-ray sequence was established, based on electron density combined with the experimentally determined N-terminal sequence, and the sequence information derived from mass spectroscopic analysis. X-ray sequence-based homolog search and the three-dimensional structure reveal that trichomaglin is a novel S-like RNase, which was confirmed by biological assay. Trichomaglin molecule contains an additional β sheet in the HV_b region, compared with the known plant RNase structures. Fourteen cystein residues form seven disulfide bridges, more than those in the other known structures of S- and S-like RNases. His43 and His105 are expected to be the catalytic acid and base, respectively. Four hydrosulfate ions are bound in the active site pocket, three of them mimicking the substrate binding sites.

Introduction

Ribonucleases (RNases) are important enzymes ubiquitous in distribution and related to RNA metabolism. RNases catalyze the degradation of RNA by hydrolyzing the phosphodiester bond of RNA, via 2′,3′ cyclic nucleotides, to 3′ mononucleotides or oligonucleotides with a 3′phosphate (Deshpande and Shankar, 2001). RNases are classified into three groups: RNase A, RNase T1, and RNase T2, with molecular masses of approximately 14, 12, and 24–36 kDa, respectively. The enzymes belonging to RNase A and RNase T1 family are mainly pyrimidine base specific and guanylic acid specific, respectively, and their optimum pH is 7.0–8.0. The enzymes in the RNase T2 family are either nonspecific or adenylic acid preferential and show pH optima in the range of 4.0–6.0 (Deshpande and Shankar, 2001).

RNase T2 is a large family of ribonuclease, widespread in various organisms, from viruses, protozoas, and fungi to plants and animals. The amino acid sequence comparison of the RNases in RNase T2 family reveals that two common sequence, CAS I and CAS II, are conserved from viruses to animals (Deshpande and Shankar, 2001).

The RNases from plant kingdom are classified into two distinct subclasses, S-RNases and S-like RNases. Gametophytic self-incompatibility is a genetically controlled fertilization system in plants, which prevents the pistil from accepting the self-pollen and suppresses the growth of the self-pollen tube (Newbigin et al., 1993; de Nettancourt, 2001). Several members of RNases from styles of higher plants in the RNase T2 family are sequence homologous and associated with gametophytic self-incompatibility, and they are called S-RNases. Some extracellular and intracellular RNases in RNase T2 family were isolated from seeds, fruits, and other tissues of self-incompatible plants and are not involved in self-incompatibility, which are called S-like RNases since the sequence comparison shows that they are homologous to S-RNases. The roles of S-like RNases in the self-incompatible plants are unknown, and further studies on their structures will provide a basis to understand the mechanism of action of S-like RNases.

The three-dimensional structure of RNase Rh, a fungal RNase from *Rhizopus nivers*, was reported in 1996, which was the first three-dimensional structure in the RNase T2 family (Kurihara et al., 1996). Four crystal structures of RNases from plants were reported since then: RNase MC1 and RNase LE are S-like RNases, from the seeds of bitter gourd (*Momordica charantia*) and the cultured cell of tomato (*Lycopersicon esculentum*), respectively (Nakagawa et al., 1999; Tanaka et al., 2000), and the pistil S_3-RNase from *Pyrus pyrifolia* and S_{F11}-RNase from *Nicotiana alata* (ornamental tobacco) are S- RNases (Matsuura et al., 2001; Ida et al., 2001).

Trichomaglin (MGL) is a novel protein isolated from root tuber of the plant Maganlin (*Trichosanthes Lepiniate*, Cucurbitaceae). MGL was isolated, purified, and crystallized in our laboratory, with molecular weight and isoelectric point of 24,673 Da and 5.8, respectively (Chen et al., 1999). The amino acid sequence of MGL was unknown, and it was thought to be a ribosome-inactivating protein since the RNA *N*-glycosidase activity was detected from the lyophilized powder of this protein. Two forms of crystals of MGL were grown and the X-ray data of them were collected (Wu et al., 2000). In this paper we present the crystal structure (form B) of MGL as well as the X-ray sequence and the biological activity

*Correspondence: xiazx@mail.sioc.ac.cn
[4]Present address: Biomolecular Structure Group, NCI-FCRDC, P.O. Box B, 7th Street, Frederick, Maryland 21702.
[5]Present address: Department of Chemistry and Biochemistry, University of California at San Diego, San Diego, California 92093.

of MGL, all of which reveal that MGL is an S-like RNase. Further experiments verified that the previously detected RNA *N*-glycosidase activity resulted from an impurity in the lyophilized powder, which possessed exceedingly high activity of RNA *N*-glycosidase and existed in minor quantity in the lyophilized powder. The mechanism of action of MGL is discussed in this paper, based on the crystal structure of MGL in which three hydrosulfate ions are bound in the active site pocket and mimic the substrate binding sites.

Results

N-Terminal Sequence

The sequence of the 20 *N*-terminal residues was determined to be DEREFDYFILALQWAGTSCR.

Overall Structure

The refinement statistics of the crystal structure are summarized in Table 1. The R factor and R_{free} are 19.9% and 23.7%, respectively. The rms deviations form the ideal bond lengths and bond angles are 0.0058 Å and 1.3°, respectively. There is 1 MGL molecule in each asymmetric unit, which consists of 209 amino acid residues; however, the 3 N-terminal residues showed weak electron density and could not be located. Each molecule contains 7 disulfide bridges, 62 water molecules, and 4 hydrosulfate ions. The ribbon diagram of the MGL molecule is shown in Figure 1.

X-Ray Sequence and Homolog Search Results

The X-ray sequence of MGL is shown in Figure 2. The N-terminal residues Glu4–Arg20 agree with the experimentally determined N-terminal sequence.

Partial sequence of MGL was confirmed with mass spectroscopic analysis. The following segments from MS match the X-ray sequence and the residue numbers in the X-ray sequence are indicated in the parentheses: EFDYFILALQWAGTSCR (4–20), ADSPTQFTI (34–42), LVEYWPTYR (75–83), CALEQSCNNR (84–93), EILWGQ QYEK (95–104), HGTCASPVIK (105–114), TLKLFMK (123–129), ALEDAGIVASNSK (135–147), DIVVAVES AVGAR (153–165), CDEEGLVQK (170–178), and LSLCFDK (179–185). The total number of the residues covered by these segments is 114.

The X-ray sequence-based homolog search reveals that MGL is homologous to three S-like RNases: AhSL28 (from *Antirrhinum majus x Antirrhinum hispanicum*, CAD33235), AmSL28 (from *Antirrhinum mollissimum*, CAD50874), and RSN2 (from *Arabidopsis thaliana*, NP_030524), with approximately 50% identical residues in their sequences. Figure 2 shows the X-ray sequence of MGL aligned with these three proteins as well as with the four proteins with known crystal structures, RNase LE, RNase MC1, S_3-RNase, and S_{F11}-RNase.

Secondary Structure

The secondary structure of MGL is shown in Figures 1 and 2. The MGL molecule contains six α helices, two 3_{10} helices, and four β sheets composed of a total of ten β strands. One of the β sheets, β_a, is composed of four β strands, with strands β_{a-1} and β_{a-4} in the middle of the sheet and strands β_{a-2} and β_{a-3} adjacent to strands β_{a-1} and β_{a-4}, respectively, and each pair of the two adjacent β strands are antiparallel to each other. The other three β sheets, β_b, β_c, and β_d, are all two-stranded antiparallel ones.

The secondary structures of MGL are very similar to those of the other four S- and S-like RNases with known crystal structures, as shown in Figure 2. However, MGL contains an additional β sheet, β_b, which is located in one of the hypervariable regions HV_b (Green, 1994; Matton et al., 1997) proposed to be the S-allele discrimination sites of S-RNases during the self-incompatibility reaction. There is no insertion or deletion in MGL in this region, compared with the three S-like RNases with unknown crystal structures, AhSL28, AmSL28, and RNS2. However, there are two to seven insertions in MGL when it is compared to RNase LE, RNase MC1, S_3-RNase, and S_{F11}-RNase, so that their secondary structures are very different form each other in this region. For example, in this region there is a 3_{10} helix and an α helix in RNase S_3 and RNase S_{F11}, respectively, both of which are S-RNases, and only the extension toward N terminus of an α helix was observed in S-like RNases LE and MC1. In the MGL molecule one of the 3_{10} helices is located in HV_a region (Green, 1994; Matton et al., 1997), where there is no deletion or insertion compared with the other four known structures. This region contains a 3_{10} helix and an α helix in RNase MC1 and RNase LE, respectively, and two N-terminal residues of an α helix in both RNase S_3 and RNase S_{f11}. Strand β_{a-4} of MGL contains Gly174-Ser180 and β_{d-1} contains Cys182-Asp184, and they belong to two different β sheets, but they were assigned to be a single β strand called β-5 in the other known crystal structures of the S- and S-like RNases.

Disulfide Bridges

The electron density of the disulfide bridges is defined very well in MGL. There are a total of 14 cysteins which form 7 disulfide bridges (Figure 1): Cys19-Cys31, Cyr25-Cys32, Cys26-Cys84, Cys57-Cys90, Cys58-Cys108, Cys170-Cys198, and Cys182-Cys192. Figure 3 shows the electron density of the first three disulfide bridges.

The disulfide pattern of MGL is shown in Figure 2, and those in RNase LE, RNase MC1, S_3-RNase, and S_{F11}-RNase are also shown for comparison. There are less disulfide bridges in the other four known S- and S-like RNase structures, six in RNase LE and four in RNase MC1, S_3-RNase, and S_{F11}-RNase. MGL shares three common disulfide bridges, Cys58-Cys108, Cys170-Cys198, and Cys182-Cys192, with all of the other four structures. There is another disulfide bridge, Cys26-Cys84, in MGL, which is common to RNase LE. In the segment Arg20-Cys32 of MGL, there are four cysteins forming two disulfide bridges, Cys19-Cys31 and Cyr25-Cys32; however, there are only two cysteins forming one disulfide in the other four structures. Therefore, the disulfide bridge pattern in this segment of MGL differs from those of the other four RNases. Cys57 and Cys90 exist only in MGL and form an additional disulfide bridge. All the 14 cysteins of MGL are well aligned with those of AhSL28,

Table 1. Statistics of Data Collection, MIR Phasing, and Crystallographic Refinement

Crystal	Native	K_2PtCl_4 Derivative	$(NH_4)_3IrCl_6$[a] Derivative
Soaking concentration (mM)		0.5	100[a]
Soaking time (hr)		20	24
Space group	$P6_1$[b]	$P6_1$[b]	$P6_1$[b]
Unit cell parameters			
a (Å)	93.99	93.69	93.70
b (Å)	93.99	93.69	93.70
c (Å)	57.77	57.59	57.83
No. of molecule per asymmetric unit	1	1	1
Resolution (Å)	2.2	2.8	3.0
No. of unique reflections	14,893	7,210	5,892
R_{sym} (%)[c]	10.0 (66.2)[d]	8.9 (37.4)[d]	13.2 (38.3)[d]
Data completeness (%)	99.8 (100.0)[d]	99.7 (99.4)[d]	99.5 (95.3)[d]
Reflections with $I > 3\sigma$ (I) (%)[e]	77.3 (41.2)[d]	86.6 (60.7)[d]	83.0 (54.0)[d]
ΔF_{iso} (%)[f]		18.6	20.7
ΔF_{ano} (%)[g]		5.9	8.3
R_{cullis}[h]		0.826	0.783
R_{kraut}[i]		0.1167	0.126
Phasing power[j]		0.900	1.029
Figure of merit		0.273	0.278
Mean figure of merit[k]		0.941	
No. of amino acid residues	209		
No. of HSO_4^-	4		
No. of solvent molecules	62		
R factor (%)[l]	19.9 (25.3)[d]		
Free R factor (%)[m]	23.7 (30.6)[d]		
Rmsd[n]			
Bond lengths (Å)	0.0058		
Bond angles (°)	1.3		
Dihedrals (°)	23.6		
Improper (°)	0.79		
Mean temperature factors (Å²)[o]			
Main chain	33		
Side chain	34		
HSO_4^-	70[p]		
Solvent	38		
Residues in Ramachandran plot (%)[q]			
In most favored region	86.8		
In additional allowed region	12.1		
Mean atomic coordinate error (Å)[r]	0.26		

[a] In order to prevent the crystals from cracking, the crystals were soaked in the reservoir solution containing $(NH_4)_3IrCl_6$ at a series of concentrations, increasing gradually from 10 to 100 mM with an interval of 10 mM, 30 min at each concentration, and 24 hr at the final one. [b] The X-ray data indicated that the space group was $P6_1$ or $P6_5$, and it was determined to be $P6_1$ during the phasing based on the anomalous dispersion. [c] $R_{sym} = \Sigma(|I - <I>|)/\Sigma(I)$. [d] The numbers in parentheses correspond to the data in the highest resolution shell (2.20–2.25, 2.80–2.87, and 3.00–3.07 Å for native, Pt, and Ir derivatives, respectively). [e] The percentage of the reflections with mean signal-to-noise ratio larger than 3. [f] Isomorphous difference ratio between each heavy atom derivative (F_{ph}) and native (F_p), calculated at 3 Å resolution. $\Delta F_{iso} = <||F_{ph}| - |F_p||>/<(|F_{ph}| + |F_p|)/2>$. [g] Anomalous difference ratio of each heavy atom derivative, calculated at 3 Å resolution. $\Delta F_{ano} = <||F_{ph}(+)| - |F_{ph}(-)||>/<(|F_{ph}(+)| + |F_{ph}(-)|)/2>$. [h] $R_{cullis} = sqrt<E^2>/sqrt<(|F_{ph}| - |F_p|)^2>$, calculated at 3 Å resolution, where E is the residual lack of closure error. [i] $R_{kraut} = \Sigma|(|F_p + F_h| - |F_{ph}|)|/\Sigma|F_p|$, calculated at 3 Å resolution, where F_h is the heavy atom structure factor amplitude. [j] Phasing power = $sqrt<F_h^2>/sqrt<E^2>$, calculated at 3 Å resolution. [k] Mean figure of merit after density modification, calculated at 3 Å resolution. [l] $R = \Sigma||Fo| - |Fc||/\Sigma|Fo|$. [m] R_{free} is calculated from 10% of the X-ray data, randomly selected for cross-validation (Brunger, 1992). [n] Root-mean-square deviation. [o] B value from Wilson plot is 19.8 Å². [p] The mean temperature factors of SO4 251, SO4 252, SO4 253, and SO4 254 are 55, 63, 81, and 82, respectively. [q] Calculated using program PROCHECK (Morris et al., 1992). [r] Obtained from Luzzati plot (Luzzati, 1952).

AmSL28, and RNS2, which suggests that MGL should share the same disulfide patterns with these three RNases.

Compared with the disulfide bridge pattern of the RNase Rh, only two disulfide bridges of MGL, Cys58-108 and Cys170-Cys198, are common to this fungal RNase.

Tertiary Structure Comparison

The tertiary structure of MGL is similar to the other four known structures of RNases from plants (Nakagawa et al., 1999; Tanaka et al., 2000; Matsuura et al., 2001; Ida et al., 2001). The most striking differences were observed in the segments Gly22-Pro37, Tyr49-Pro55, and Ala85-Asn95. The segment Gly22-Pro37 of MGL contains three to eight insertions compared with the other four known structures, leading to large difference in this surface loop. Tyr49-Pro55 is a flexible surface loop, and the main chain conformations of this segment obviously differ from each other in all of the five structures although there are only one or two insertions or deletions in the sequence. The segment Ala85-Asn95 of MGL is located

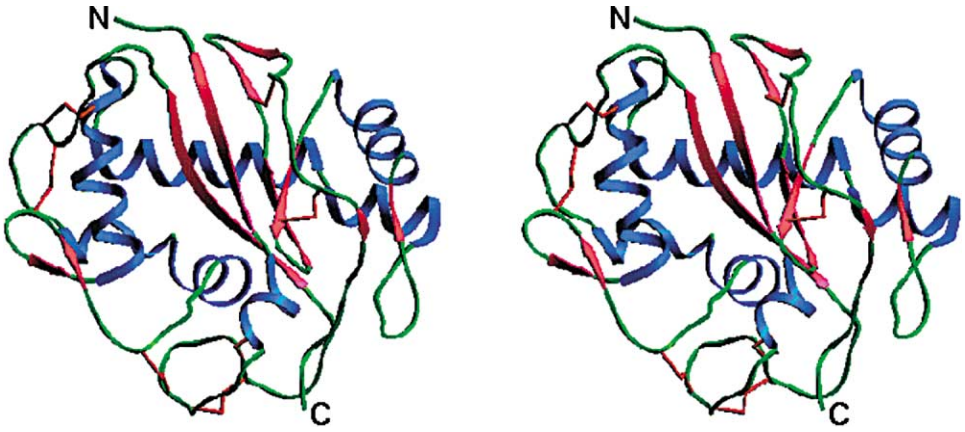

Figure 1. Ribbon Diagram of MGL

Helices, β sheets, and random coils are shown in blue, red, and green, respectively. Disulfide bridges are shown in orange. The N and C termini are labeled. This diagram was prepared using the program SETOR (Evans, 1993).

in the HV_b region and contains an additional β sheet, as described above. The differences in local conformation can be observed for some other surface loops where there is insertion or deletion in the sequence. Based on the sequence alignment, the hypervariable region HV_a of MGL is in the segment Asp59-Leu75, and the main

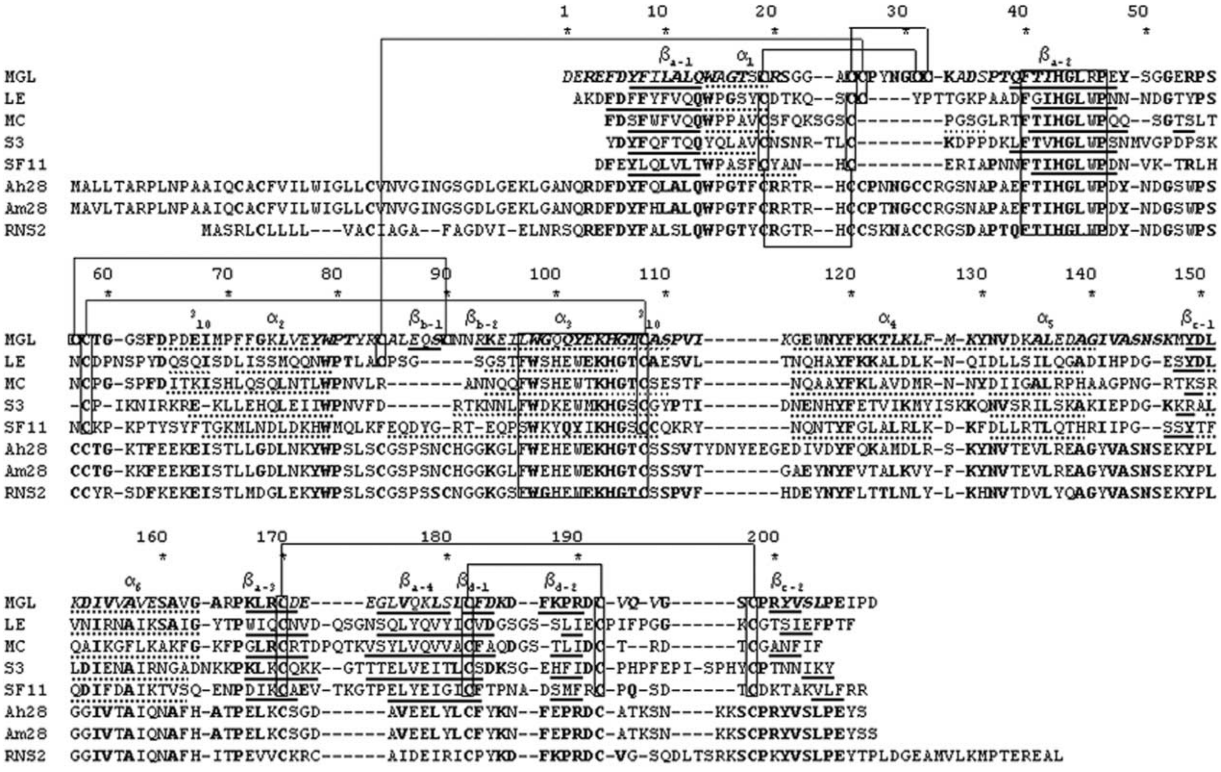

Figure 2. X-Ray Sequence of MGL

X-ray sequence of MGL, aligned with those of RNase LE from the cultured cell of tomato (*Lycopersicon esculentum*) (Jost et al. 1991), RNase MC1 from the seeds of bitter gourd (*Momordica charantia*) (Ide et al.,1991), the pistil S_3-RNase from *Pyrus pyrifolia* (Ishimizu, 1998), S_{F11}-RNase *Nicotiana alata* (ornamental tobacco) (Kheyr-Pour et al., 1990), AhS28 (from *Antirrhinum majus x Antirrhinum hispanicum*, CAD33235), AmSL28 (from *Antirrhinum mollissimum*, CAD50874), and RSN2 (from *Arabidopsis thaliana*, NP_030524), which are abbreviated to LE, MC, S3, SF11, Ah28, Am28, and RNS2, respectively, in this diagram. The residue numbers are based on the X-ray sequence of MGL. The β sheets and helices of the first five proteins are underlined with solid lines and dotted lines, respectively, and α helices, 3_{10} helices, and β sheets of MGL are labeled. Disulfide bridges in the first five structures are also indicated. The residues in the other seven RNases that are identical to those in MGL are shown in bold, and the corresponding residues of MGL are in bold as well. The residues matching MS sequence information and/or with the determined N-terminal sequence are shown in italics. Two boxes denote CAS I and CAS II.

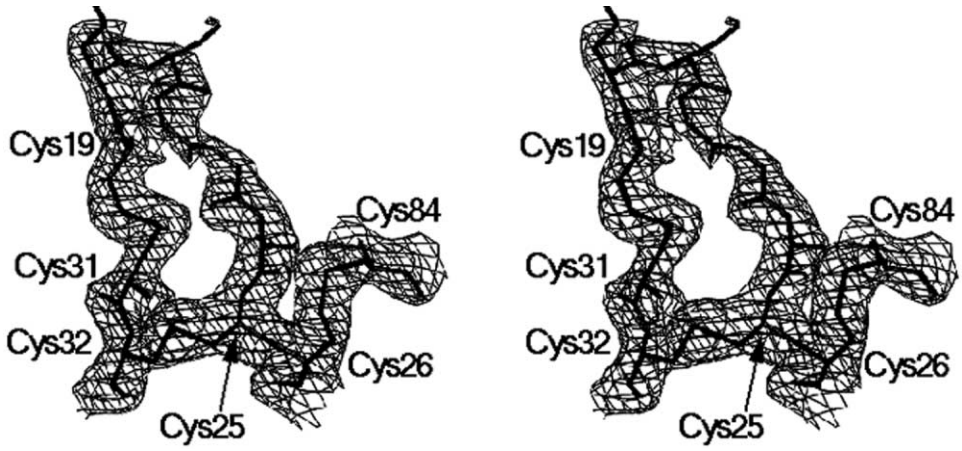

Figure 3. (2Fo-Fc) Electron Density of Three Disulfide Bridges, Cys19-Cys31, Cyr25-Cys32, and Cys26-Cys84, Contoured at 1.0 σ
The six cysteins are labeled. This diagram was prepared using the program TURBO-FRODO.

chain conformation of this segment is similar to the other four structures. Taken as an example, the tertiary structure of MGL compared with RNase LE is shown in Figure 4, and the rms deviations for the Cα atoms between the two structures is 2.7 Å.

Catalytic Sites

The fungal RNase Rh is the well-studied member in RNase T2 family, and chemical modification, site-directed mutagenesis, and crystal structure revealed that two histidines and one glutamic acid construct its catalytic sites (Kurihara et al., 1996; Sanda et al., 1985a, 1985b; Ohgi et al., 1993). The catalytic sites are well conserved in the primary structures of all the S- and S-like RNases when their sequence are aligned with that of RNase Rh except that the glutamic acid is replaced by glutamine in S_{F11}-RNase. For example, the catalytic sites of RNase LE are His39 (catalytic acid), His97 (catalytic base), and Glu93, equivalent to His46, His109, and Glu105 of RNase Rh.

Based on the sequence alignment shown in Figure 2, His43 and His105 of MGL are equivalent to His39 and His97 of RNase LE, and they are expected to be the catalytic acid and base for the catalysis. It was reported that Glu93 and Lys96 of RNase LE also played important roles in the catalysis, and the latter is a conserved residue (Lys104 in MGL); however, the former is replaced by a glutamine (Gln101) in MGL as in S_{F11}-RNase (Kheyr-Pour et al., 1990). The electron density of His43, His105, Gln101, and Lys104 of MGL is shown in Figure 5. The orientations of the imidazole ring of His105 and its equivalent ones in the other four known structures are similar to each other, and the atoms NE2 deviate from each other slightly when they are superimposed. However, the orientation of the imidazole ring of His43 in MGL is similar to those in S_3-RNase and S_{F11}-RNase and approximately 90° different from those in RNase LE and RNase MC1, and the atoms NE2 in the five structures exhibit larger deviation. The conformation of Gln101 side chain is very similar to the equivalent residues in the other four known structures except that the one in S_3-RNase deviates obviously.

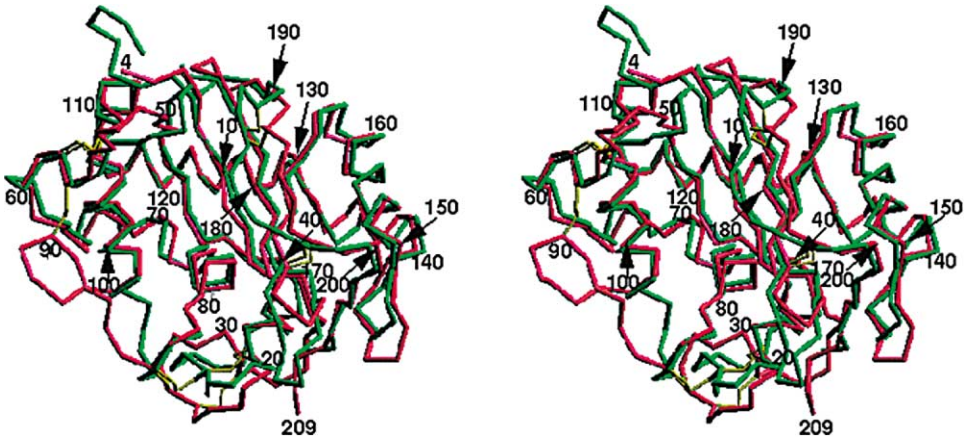

Figure 4. Cα Backbone of MGL Superimposed with that of RNase LE
Cα backbone of MGL shown in red is superimposed with that of RNase LE shown in green. Disulfide bridges are shown in yellow. Every tenth Cα atoms of MGL are labeled. Asp4 and Asp209 are also labeled. This diagram was prepared using the program SETOR.

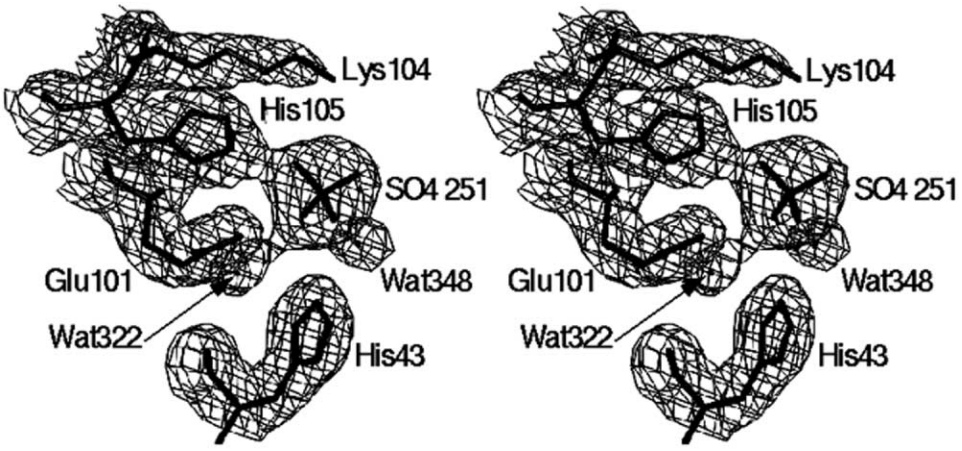

Figure 5. The Catalytic Sites His43 and His105 Are Superimposed with the (2Fo-Fc) Electron Density Contoured at 1.0 σ

Gln101, Lys104, and two water molecules as well as the hydrosulfate ion SO4 251 are also shown. This diagram was prepared using the program TURBO-FRODO.

Hydrosulfate Ions

The MGL crystals (form B) were grown in the presence of 0.15 M $KHSO_4$. Four hydrosulfate ions were found in the crystal structure of MGL, named SO4 251-SO4 254, as shown in Figure 6. The electron density of SO4 251 is shown in Figure 5. Three of the four HSO_4^- ions are located in the active site pocket, with SO4 251 in the center, and the distances from SO4 251 to SO4 252 and to SO4 253 are 8.7 and 9.3 Å, respectively. SO4 254 is located at the outer edge of the pocket and far from the other three HSO_4^- ions (larger than 15 Å). The four HSO_4^- ions make electrostatic and hydrogen bonding interactions with the protein atoms and water molecules, which are listed in Table 2.

RNase Activity

The X-ray sequence and the three-dimensional structure reveal that trichomaglin is a novel S-like RNase, which was verified by the biological assay for RNase activity. Figure 7 shows the variation of the difference optical density at 260 nm of MGL with the change in pH value, indicating that MGL catalyzes the hydrolysis of RNA, and the RNase activity of MGL is pH dependent, with an optimum pH of 4.0.

Discussion

The Reliability of the X-Ray Sequence of MGL

The X-ray-based sequence of a protein is usually of inherently insufficient accuracy because some of the residues can not be distinguished, such as between Asp and Asn, and Glu and Gln, and the side chains of the residues at molecular surface may display weak electron density. However, the X-ray sequence of trichomaglin reported in this paper was obtained by combining with the partial sequence information from mass spectroscopic analysis. A total of 114 amino acid residues of the X-ray sequence agree with the sequence information from MS and/or the experimentally determined N-terminal sequence, and Asp1-Arg3 are also determined

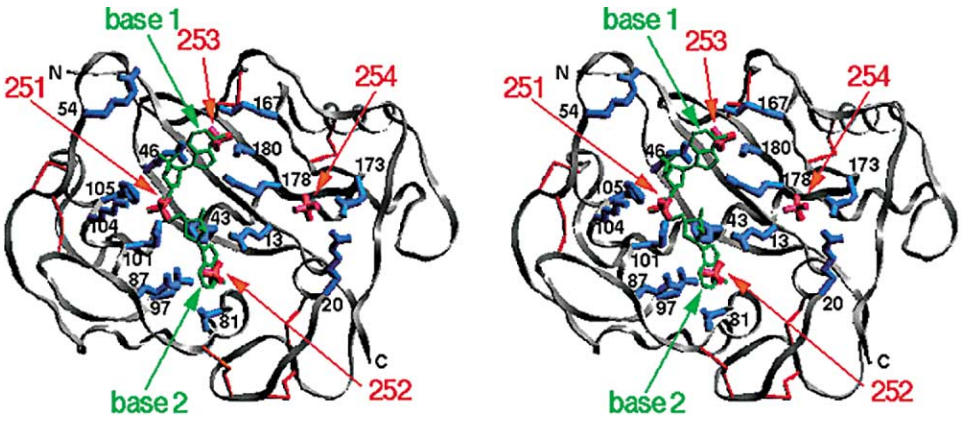

Figure 6. Hydrosulfate Ions and the Modeled Dinucleotide Substrate

The backbone of MGL is shown in gray and the disulfide bridges in orange. The side chains interacting with the substrate or hydrosulfate ions and the side chain of Arg54 are shown in blue and labeled, hydrosulfate ions in red and labeled as 251-254, and the dinucleotide model in green with the two bases labeled. The N and C termini are also labeled. This diagram was prepared using the program SETOR.

Table 2. Interactions of Hydrosulfate Ions with MGL

Hydrosulfate Ion	Atom 1 in HSO_4^-	Atom 2 in MGL	Distance (Å)
HSO_4^- 251	O1	Lys104 NZ	3.37
	O2	His105 NE2	2.65
	O2	Gln101 NE2	3.10
	O3	His43 NE2	2.79
	O3	Gln101 NE2	3.33
	O3	Wat348	2.98
	O4	Wat332	3.44
HSO_4^- 252	O1	Gln13 OE1	3.06
	O1	Gln13 NE2	3.13
	O2	Tyr82 N	3.12
	O3	Tyr82 N	2.92
	O3	Thr81 OG1	3.30
	O3	Wat352	2.95
	O4	Glu87 OE1	2.66
	O4	Arg83 NH2	3.18
	O4	Arg83 NE	2.67
HSO_4^- 253	O2	Ser180 OG	3.10
	O3	Wat321	3.31
	O4	Arg46 NH2	3.15
	O4	Lys167 NZ	2.79
HSO_4^- 254	O2	Glu173 OE1	2.64
	O2	Glu173 OE2	2.94
	O3	Glu172 N	3.36
	O3	Glu173OE1	3.18
	O4	Arg20 NH2	2.86
	O4	Glu173 OE1	3.37

by experiments, although they could not be located by X-ray analysis. The number of these residues comes to 117, which is 56% of the X-ray sequence shown in Figure 2. The 14 cysteins are well aligned with those in Ah28, Am28, and RNS2, and they are verified by the strong electron density characteristic of disulfide bridges formed by them. Six of the fourteen cysteins were detected by MS, and the eight remaining cysteins are approximately 4% of the X-ray sequence. Totally 60% of the X-ray sequence is thus demonstrated to be highly reliable.

Three Hydrosulfate Ions Mimic the Substrate Binding Sites

According to nomenclature of the "subsites" of RNase A (Richards and Wyckoff, 1971), P1 site was named for the binding site of 3′,5′-phosphodiester of the substrate, and a sulfate ion occupied this position in the crystal structures of bovine pancreatic ribonuclease complexed with dinucleotides, in various degrees dependent on pH. The phosphate of a cyclic substrate of pentacovalent intermediate was proposed to bind at P1 site. B1 and B2 were named for the binding sites of the first and second bases of the nucleotides linked to the phosphodiester at O3′ and O5′, respectively.

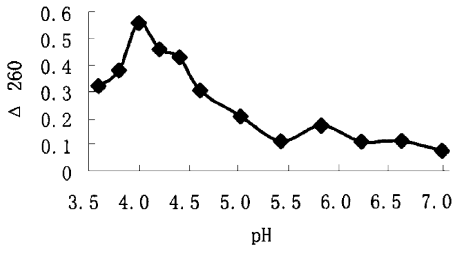

Figure 7. The Variation Curve of the Difference Optical Density, Δ 260, of MGL with pH Value

There are two water molecules located at P1 site of RNase LE (Tanaka et al., 2000), one forming hydrogen bonds with side chains of His92 and Lys96 and the other with Glu93 and His97, and these water molecules are conserved in RNase Rh. The crystal structure of RNase Rh complexed with inhibitor 2′-AMP showed that the P1 site was occupied by the 2′-phosphate group (Nakamura et al., 1993). In MGL structure, the negatively charged SO4 251 was found at P1 site, forming electrostatic interactions with the positively charged side chains of the active site residues His43, His105, and Lys104, and forming hydrogen bonds with Gln101 and two water molecules (Table 2; Figure 5). The position of this HSO_4^- ion well agrees with the two conserved water molecules found in the P1 site of RNase LE and RNase Rh. It can be deduced that SO4 251 mimics the binding site P1 of the 3′,5′-phosphodiester group of the substrate RNA.

The mutagenesis studies suggested that the hydrophobic pocket containing the aromatic side chains of Trp49 and Tyr57 in RNase Rh was the base binding site B1 of the substrate, and the adenine base of 2′-AMP in the complex was stacked between Trp49 and Tyr57 of RNase Rh (Nakamura et al., 1993). These two residues correspond to Trp42 and Tyr50 in RNase LE (Tanaka et al., 2000). The tryptophan is highly conserved in S- and S-like RNases, and the tyrosine is replaced by tryptophan in RNases AhSL28, AmSL28, and RNS2 and by nonaromatic residues in RNase MC1, S_3-RNase, and S_{F11}-RNase (Figure 2). However, these two residues are replaced in MGL by a pair of arginines, Arg46 and Arg54. Figure 8 shows the electron density of Arg46, which is the only nontryptophan residue at this position in the so-called CAS I region (Deshpande and Shankar, 2001; Tanaka et al., 2000) of S- and S-like RNases (Figure 2). The distance between these two arginine side chains is approximately 8 Å, similar to that between the two rings of Trp42 and Tyr50 in RNase LE. However, each of Arg46 and Arg54 in MGL points to different direction from that of Trp42 or Tyr50 in RNase LE, and the two arginines in MGL are 2–3 Å farther from P1 site than Trp42 and Tyr50 from P1 in RNase LE. Therefore, it is too far from P1 to locate the base 1 between Arg46 and Arg54. SO4 253 forms electrostatic interactions with the positively charged side chains of Arg46 and Lys167, and it is hydrogen bonded with Ser180 and a water molecule. The property of the electrostatic interactions probably suggests that this hydrosulfate ion might mimic a phosphodiester group of the substrate. However, this hydrosulfate ion is located 9.3 Å away from the P1 site, too far to locate another phosphodiester group of the substrate. A model of a dinucleotide substrate containing two adenylic acid residues was manually built, in which the phophodiester group is at P1 site, and the first base, base 1, can be modeled at the site of SO4 253 (Figure 6), suggesting that SO4 253 is likely to mimic the B1 site. In addition, if the first adenylic acid residue extends from its 5′-phosphate, the base of the preceding nucleotide residue in RNA can be stacked between Arg46 and Arg54.

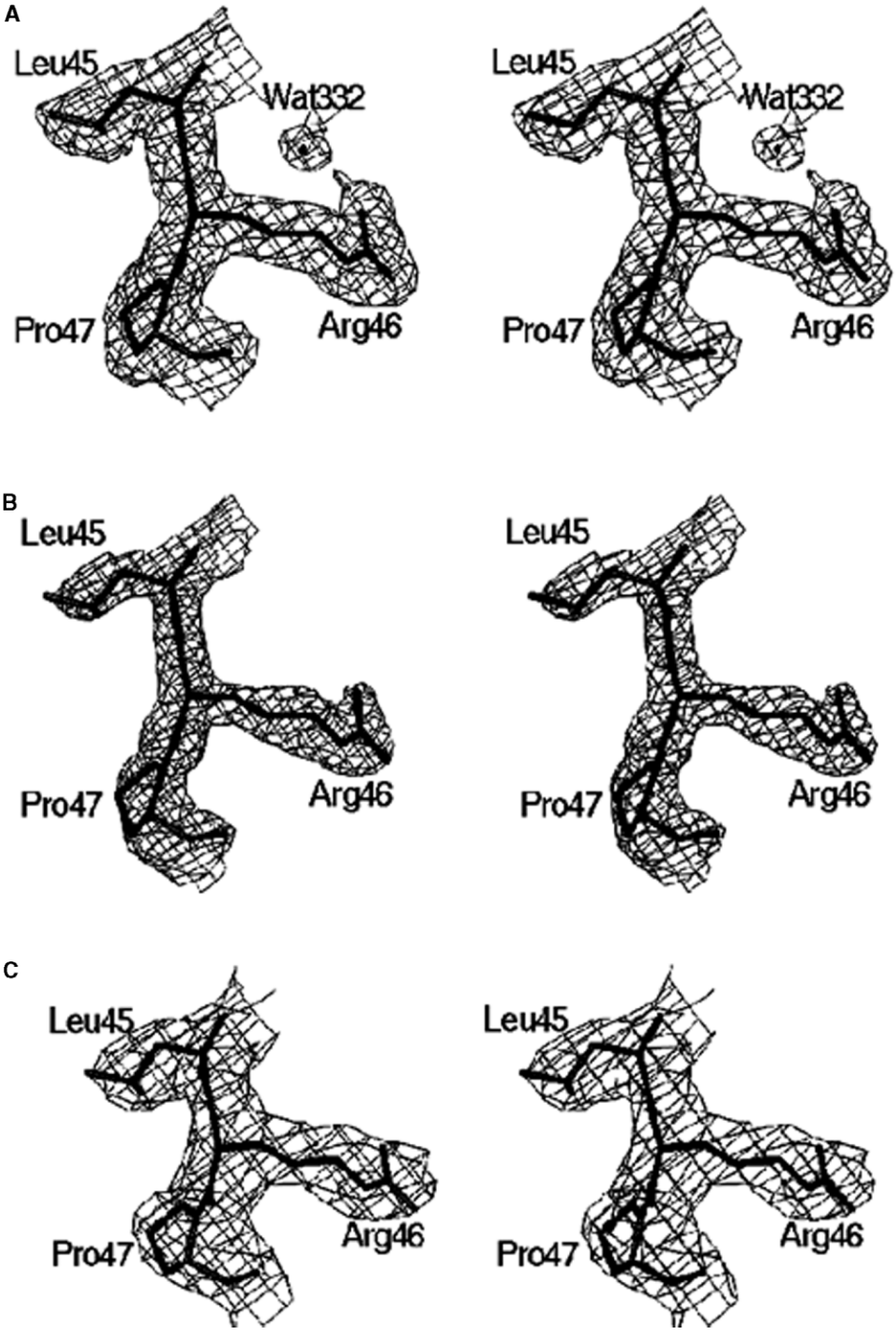

Figure 8. The Electron Density of Arg46

(A) (2Fo-Fc) electron density, contoured at 1.0 σ.

(B) The simulated annealing (Fo-Fc) electron density, contoured at 3.0 σ.

(C) The MIR map, contoured at 1.0 σ. These diagrams were prepared using the program TURBO-FRODO.

The hydrosulfate ion SO4 252 is hydrogen bonded to the side chains of Gln13, Thr81, and Glu87, and it forms electrostatic interactions with Arg83. Gln13 of MGL corresponds to one of the B2 site residues, Gln32 of RNase Rh, and it is conserved in S- and S-like RNases except for S_{F11}-RNase. The crystal structure of RNase Rh/d(ApC) complex showed that the cytocine base of the dinucleoside phosphate d(ApC) was stacked with the side chain of Phe101 of RNase Rh (Hamashima, 1994), which corresponds to the putative B2 site, Phe89 of

RNase LE (Tanaka et al., 2000), and is conserved in most of the S- and S-like RNases. However, this phenylalanine is replaced by Leu97 in MGL and by a serine residue in S_{F11}-RNase. The position of Leu97 of MGL is similar to that of Phe89 of RNase LE, but the side chain of the former is smaller than the latter and SO4 252 is approximately 4 Å from Leu97 side chain. The second base, base 2, can be modeled at the site of SO4 252 (Figure 6), similar to the putative B2 site of RNase LE, which suggests that SO4 252 is likely to mimic the B2 site.

In the model shown in Figure 6 there are a few close contacts, such as base 1 with the guanidino group of Arg46, and base 2 with Arg83 (approximately 2 Å), and these side chains are very flexible and there are sufficient spaces around them to allow their moving away when the substrate enters the pocket. A mechanism of the 2′,3′-cyclizing RNase was previously proposed (Deshpande and Shankar, 2001; Deavin et al., 1966), including a transphosphorylation step and a hydrolysis step. Our model well agrees with this mechanism. O5′ of the second ribose can form a hydrogen bond of 2.9 Å with the atom NE2 of His43, and His43 can act as a general acid catalyst to donate a proton. The atom NE2 of His105 is 3.2 and 3.6 Å from O3′ and O_2′ of the first ribose, respectively, and His105 can act as a general catalytic base to remove the hydrogen at 2′-OH. The positively charged Lys104 is 4.1 Å from the phosphorous and 3.0 Å from the closest oxygen of the phosphate, and it may help the polarization of the P-O bond. Gln101, which is 3.9 and 3.4 Å from the phosphorous and the oxygen, may help stabilize the polarized P-O bond through hydrogen bonds. Then O2′ nucleophilically attacks the polarized phosphorous, leading to a pentacovalent intermediate, and the phosphodiester bond is cleaved. In the second step, the hydrolysis of the intermediate takes place to form a 3′-nucleotide.

Biological Activity

The crystal structure of MGL and its X-ray sequence aligned with other S- and S-like RNases reveal that MGL is an S-like RNase. The biological assay results show that MGL catalyzes the hydrolysis of yeast RNA, confirming the RNase activity of MGL. However, in the previous study the RNA *N*-glycosidase activity was detected from the lyophilized powder of MGL so that it was thought to be a ribosome-inactivating protein (Chen et al., 1999; Wu et al., 2000). After the structure of MGL was solved, further experiments using equal amount of the dissolved single crystals and the lyophilized powder of MGL verified that the detected RNA *N*-glycosidase activity resulted from an impurity in the lyophilized powder. This impurity possessed exceedingly high activity of RNA *N*-glycosidase, and it existed in minor quantity in the lyophilized powder but not in the crystals. Here we rename this minor component to be β-trichomaglin, and the name trichomaglin is used for the major component of which the crystal structure and X-ray sequence are reported in this paper.

Experimental Procedures

N-Terminal Sequence Determination

The MGL crystals were washed using the reservoir solution, dissolved in water, concentrated at low temperature using Centricon 10, and lyophilized. The N-terminal sequence was determined using Edman degradation method on an amino acid sequencer.

Protein Sequence Analysis by On-Line LC-MS/MS

The lyophilized MGL was dissolved in water at 0.5 mg/l. The protein was reduced with 10 mM DTT and then alkylated with 20 mM iodoacteamide followed by dialysis against 50 mM NH_4HCO_3 to remove excessive reagents before tryptic digestion. The digestion was carried out at 1:50 (E/S) of trypsin (Boehringer Mannheim, UK) to protein for 18 hr at 37°C.

An aliquot of the digest (800 fmol) was injected for automated online LC-MS/MS analysis. The peptides were separated on a 75 μm id × 15 cm I PepMap C18 column using an Ultimate LC system (LC Packings, Netherlands) delivering a gradient 0%–48% acetonitrile (containing 0.05% formic acid) in an hour with the flow rate at 200 nl/min. The eluting peptides were directly infused to a QTOF hybrid mass spectrometer (Micromass, UK) fitted with a nano-flow Z-spray electrospray source. The instrument was operated at the data-dependent acquisition mode, and it selected up to three coeluting multiply charged ions (2+ and 3+) with the intensity above the set threshold for peptide sequencing by tandem MS. The LC-MS/MS data were processed with Masslynx 3.4 software (Micromass, UK) and submitted to Mascot search (Matrix Science, UK) against nonredundant protein data base including all species. The unassigned quality MS/MS spectra were manually interpreted to get the peptide sequence.

MIR Phasing

The phase problem was solved by multiple-isomorphous replacement (MIR) method. After a comprehensive search, two heavy atom derivatives, prepared by soaking in K_2PtCl_4 and $(NH_4)_3IrCl_6$, respectively, were finally selected for phasing, and the soaking conditions of them are listed in Table 1. The X-ray data were collected on a MAR Research 300 imaging plate detector system, and the data were processed using the program DENZO and SCALEPACK (Otwinowski and Minor, 1997). The data collection statistics of the native MGL and the two derivatives are also shown in Table 1.

The Patterson functions were solved at 3 Å resolution, based on the space group $P6_1$, using program SOLVE (Terwilliger et al., 1987), giving six Pt and seven Ir atom positions and an overall Z score of 28.8. The heavy atom parameters were further refined, and the phases were computed and further improved by density modification with solvent flipping protocal, using program CNS (Brunger et al., 1998). The electron density map at 3 Å resolution was computed. The phasing statistics are shown in Table 1. The overall Z-score based on the space group $P6_5$ was 8.44, leading to a much worse electron density map.

Model Building and Refinement

The polypeptide chain was traced and the initial model was built, based on the electron density map at 3 Å resolution with the MIR phases. The initial X-ray sequence was established, and the experimentally determined N-terminal sequence was used to identify the N terminus of the polypeptide chain. Then, the crystallographic refinement was carried out by refining the atomic positions iteratively with simulated annealing procedure followed by temperature factor refinement. The manual refitting was carried out based on the (2Fo-Fc) and (Fo-Fc) electron density maps. The X-ray data were extended gradually up to 2.2 Å resolution and the X-ray sequence was improved during the refinement. The solvent molecules were added into the model in the late stage and only those with $B < 50$ Å^2 are included in the final model. At this point the X-ray sequence was checked with the partial sequence information from mass spectroscopic analysis, and then the sequence homolog search was carried out based on the X-ray sequence, using program BLAST (Altschul et al., 1997). The model was further refined and the X-ray sequence was further improved. The structure was checked with the simulated annealing "omit" maps, which were computed by systematically omitting ten residues at a time with one residue overlap between two adjacent "omit" maps. In the late stage, more MS sequencing information were available, and the X-ray sequence was further improved to match both the MS sequence and the electron density.

The refinement was carried out using program CNS and the man-

ual fitting was performed using programs TURBO-FRODO (Roussel and Cambillau, 1991) and O (Jones et al., 1991).

Biological Assay

MGL was assayed by measuring the optical density at 260 nm of the soluble nucleotide fragments which were released by the hydrolysis of 1 mM yeast RNA (purchased from Dong Feng Biochemical Technology Company, Shanghai, China) with 0.003 mg MGL, at 37°C for 30 min, in acetate or phosphate buffers (pH range 3.6–7.0) containing 0.1 N NaCl. The reaction was stopped by adding $HClO_4$. The solution was centrifuged at 10,000 rpm for 10 min, and 50 μl of supernatant was diluted with 0.1 N HCl into 250 μl. The optical density was measured on a Beckman DU spectrophotometer. The reference optical density was measured when the reaction was terminated immediately after the RNA was added by adding 20 μl 25% $HClO_4$, and followed by cooling in ice-bath for 10 min. The obtained difference optical density at 260 nm indicated the RNase activity of MGL.

Acknowledgments

This work was supported by two grants from National Natural Science Foundation of China (grant No. 29773059, 20173071). We are grateful to Prof. Li-Wen Niu, Prof. Mai-Kun Teng, and Dr. Qin-Qiu Huang of The University of Science and Technology of China for their support with the X-ray data collection. We also thank Dr. Andrew Cronshaw of The University of Edinburgh, Prof. Jun Zhou of Kunming Institute of Botany, and Prof. Shan-Wei Jin and Prof. Hai-Bao Chen of Shanghai Institute of Organic Chemistry for their helpful discussion.

Received: January 20, 2004
Revised: February 26, 2004
Accepted: March 27, 2004
Published: June 8, 2004

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipma, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Brunger, A.T. (1992). Free R value: a novel statistical quality for assessing the accuracy of crystal structures. Nature *335*, 472–475.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Gross-Kunstleve, R.W., Jinag, J.-S., Kuszewski, J., Nilges, N., Pannu, N.S., et al. (1998). Crystallography and NMR systems (CNS): a new software system for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

Chen, R., Xu, Y.Z., Wu, J., Pu, Z., Jin, S.W., Liu, W.Y., and Xia, Z.X. (1999). Purification and characterization of trichomaglin—a novel ribosome-inactivating protein with abortifacient activity. Biochem. Mol. Biol. Int. *47*, 185–193.

Deavin, A., Mathias, A.P., and Rabin, B.R. (1966). Mechanism of action of bovine pancreatic ribonucleoase. Nature *211*, 252–255.

de Nettancourt, D. (2001). Incompatibility and Incongruity in Wild and Cultivated Plants, Second Edition (Heidelberg: Springer-Verlag,).

Deshpande, R.A., and Shankar, V. (2001). Ribonucleases from T2 family. Crit. Rev. Microbiol. *28*, 79–122.

Evans, S.V. (1993). SETOR: hardware lighted three-dimensional solid model representations of macromlecules. J. Mol. Graph. *11*, 134–138.

Green, P.J. (1994). The ribonucleases of higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. *45*, 421–445.

Hamashima, M. (1994). The crystal structures of recombinant ribonuclease Rh and its mutants. MS thesis, Nogaoka University of Technology, Nogaoka, Japan.

Ida, K., Norioka, S., Yamamoto, M., Kumasaka, T., Yamashita, E., Newbigin, E., Clarke, A.E., Sakiyama, F., and Sato, M. (2001). The 1.55 Å resolution structure of *Nicotiana alata* S_{F11}-RNase associated with gametophytic self-incompatibility. J. Mol. Biol. *314*, 103–112.

Ide, H., Kimura, M., Arai, M., and Funatsu, G. (1991). The complete amino acid sequence of ribonuclease from the seeds of bitter gourd (*Momordica charanti*). FEBS Lett. *284*, 161–164.

Ishimizu, T., Shinkawa, T., Sakiyama, F., and Norioka, S. (1998). Primary structural features of rosaceous S-RNase associated with gametophytic self-incompatibility. Plant Mol. Biol. *37*, 931–941.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. D Biol. Crystallogr. *47*, 110–119.

Jost, W., Bak, H., Glund, K., Terpstra, P., and Beintema, J.J. (1991). Amino acid sequences of an extracellular, phosphate-starvation-induced ribonuclease from cultured tomato (*Lycopersicon esculentum)* cells. Eur. J. Biochem. *198*, 1–6.

Kheyr-Pour, A., Bintrim, S.B., Ioerger, T., Remy, R., Hammond, S.A., and Kao, T.H. (1990). Sequence diversity of pistil S-proteins associated with gametophytic self-incompatibility in *Nicitiana alata*. Sex. Plant. Reprod. 3, 88–97.

Kurihara, H., Nonaka, T., Mitsui, Y., Ohgi, K., Irie, M., and Nakamura, K.T. (1996). The crystal structure of Ribonuclease Rh from Rhizopus nivers at 2.0Å resolution. J. Mol. Biol. *255*, 310–320.

Luzzati, P.V. (1952). Traitment statistique des erreurs dans la determination des structures cristallines. Acta Crystallogr. *5*, 802–810.

Matsuura, T., Sakai, H., Unno, M., Ida, K., Sato, M., Sakiyama, F., and Norioka, S. (2001). Crystal structure at 1.5 Å resolution of *Pyrus pyrifolia* pistal ribonuclease responsible for gametophytic self-incompatibility. J. Biol. Chem. *276*, 45261–45269.

Matton, D.P., Maes, O., Laublin, G., Xike, C., Bertrand, C., Morse, D., and Cappadocia, M. (1997). Herpervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. Plant Cell *9*, 1766–1767.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. Proteins *12*, 345–364.

Nakagawa, A., Tanaka, I., Sakai, R., Nakashima, T., Funatsu, G., and Kimura, M. (1999). Crystal structure of a ribonuclease from the seeds of bitter gourd (*Momordica charantia*) at 1.75 Å resolution. Biochim. Biophys. Acta *1433*, 253–260.

Nakamura, K.T., Ishikawa, N., Hamashima, M., Kurihara, H., Nonaka, T., Mitsui, Y., Ohgi, K., and Irie, M. (1993). Protein nucleic acid recognition in ribonuclease Rh-2′-adenylic acid complex. In Third International Meeting, Ribonuleases, Chemistry, Biology, Biotechnology, *5*, Capri, Italy.

Newbigin, E., Anderson, M.A., and Clarke, A.E. (1993). Gametophytic self-incompatibility systems. Plant Cell *5*, 1315–1324.

Ohgi, K., Horiuchi, H., Watanabe, H., Iwama, M., Takagi, M., and Irie, M. (1993). Role of Asp51 and Glu105 in the enzymatic activity of a ribonuclease from *Rhizopus sp*. J. Biochem. (Tokyo) *113*, 219–224.

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation methods. Methods Enzymol. *276*, 307–326.

Richards, F.M., and Wyckoff, H.W. (1971). Bovine pancreatic ribonuclease. In The Enzymes, Volume 4, P.D. Boyer, ed. (New York: Academic Press), pp. 647–806.

Roussel, A., and Cambillau, C. (1991). TURBO-FRODO, Silicon Graphics Geometry Partners Dictionary (Mountain View, CA: Silicon Graphics).

Sanda, A., Takizawa, Y., and Irie, M. (1985a). Carboxymethylation of a ribonuclease from *Rhizopus sp*. Chem. Pharm. Bull. *33*, 4515–4521.

Sanda, A., Takizawa, Y., Iwama, M., and Irie, M. (1985b). Modification of a ribonuclease from *Rhizopus sp*. with 1-cyclohexyl-3-(2-morpholinyl-(4)-ethyl) carbodiimide-*p*-toluenesulfonate. J. Biochem. *98*, 125–132.

Tanaka, N., Arai, J., Inokuchi, N., Koyama, T., Ohgi, K., Irie, M., and Nakamura, K.T. (2000). Crystal structure of a plant ribonuclease, RNase LE. J. Mol. Biol. *298*, 859–873.

Terwilliger, T.C., Kim, S.-h., and Eisenburg, D. (1987). Generalized method of determining heavy-atom positions using the difference Patterson function. Acta Crystallogr. A *43*, 1–5.

Wu, J., Gan, J.-H., and Xia, Z.-X. (2000). Crystallization and preliminary crystallographic studies of trichomaglin, a novel ribosome-inactivating protein. Acta Crystallogr. D *56*, 1466–1467.

Accession Numbers

Atomic coordinates of MGL have been deposited in the Protein Data Bank with PDB entry ID 1SGL.